(2)

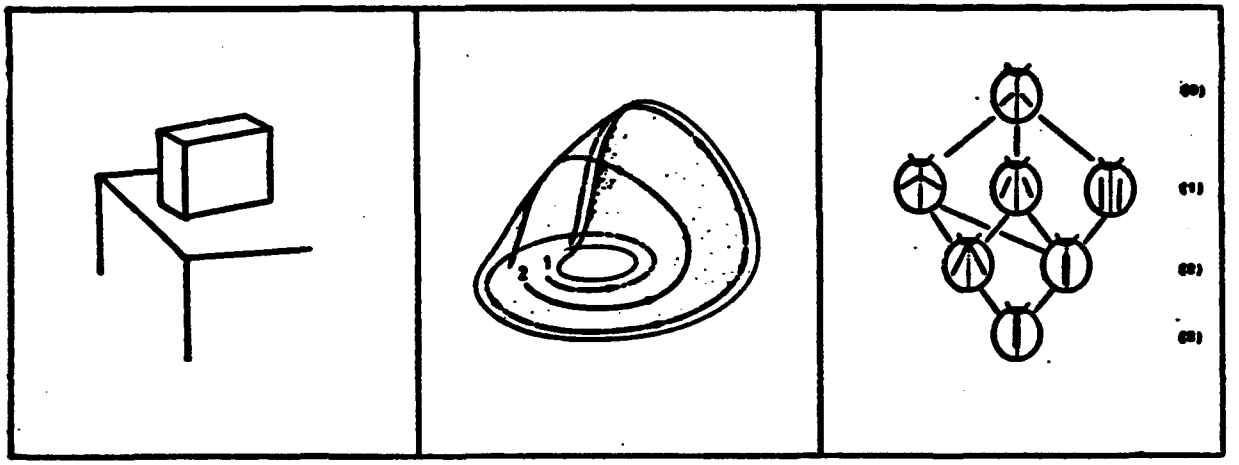


# Top-Down Influences on Bottom-Up Processing
## Final Report
## 1990 – 1993

AFOSR-TR- 93 0893

**Investigator:**
**(617) 253-5776**
**WHIT@AI.MIT.EDU**

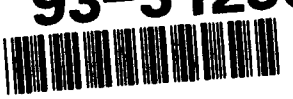
**Whitman Richards E10-120**
**Massachusetts Institute of Technology**
**Cambridge, MA 02139**



**Abstract:** Although integrated computational and psychophysical studies have considerably advanced our understanding of early visual processing (up to Marr's 2 1/2 D Sketch), much less research, by comparison, is being conducted on intermediate and higher-level vision. One reason for the scarcity is that high-level vision includes goal-directed, context-sensitive, top-down knowledge. However, little is known about these aspects of vision and how they are organized in the visual knowledge bases (for example how default preferences and categorical states are organized and related). Hence plausible, detailed computational models can't be formulated. These studies begin to reveal the structure of some aspects of cognitive visual knowledge.

Key words: Perception, key features, Bayesian models, dynamical systems, psychophysics, neural models.





October 1993

SECURITY CLASSIFICATION OF THIS PAGE

# REPORT DOCUMENTATION PAGE

| | |
|---|---|
| 1a. REPORT SECURITY CLASSIFICATION<br>Unclassified | 1b RESTRICTIVE MARKINGS |
| 2a. SECURITY CLASSIFICATION AUTHORITY | 3. DISTRIBUTION/AVAILABILITY OF REPORT<br>Approved for public release: distribution unlimited |
| 2b. DECLASSIFICATION/DOWNGRADING SCHEDULE | |
| 4. PERFORMING ORGANIZATION REPORT NUMBER(S) | 5. MONITORING ORGANIZATION REPORT NUMBER(S) |
| 6a. NAME OF PERFORMING ORGANIZATION<br>Mass. Institute of Technology<br>6b OFFICE SYMBOL (If applicable) | 7a. NAME OF MONITORING ORGANIZATION<br>Air Force Office of Scientific Res. N/L |
| 6c. ADDRESS (City, State, and ZIP Code)<br>Dept. of Brain & Cognitive Sciences<br>79 Amherst St. E10-120<br>Cambridge, MA 02139 | 7b. ADDRESS (City, State, and ZIP Code)<br>Bldg. 410<br>Bolling Air Force Base<br>Washington, DC 20332-6448 |
| 8a. NAME OF FUNDING/SPONSORING ORGANIZATION<br>AFOSR<br>8b. OFFICE SYMBOL (If applicable)<br>NL | 9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER<br>AFOSR<br>89-0504 |
| 8c. ADDRESS (City, State, and ZIP Code)<br>Same as 7b | 10. SOURCE OF FUNDING NUMBERS<br>PROGRAM ELEMENT NO. 61102F<br>PROJECT NO. 2313<br>TASK NO A9<br>WORK UNIT ACCESSION NO. |
| 11. TITLE (Include Security Classification)<br>"Top Down" Influences on "Botton-Up" Processing | |
| 12. PERSONAL AUTHOR(S) Whitman Richards | |
| 13a. TYPE OF REPORT<br>FINAL<br>13b. TIME COVERED<br>FROM 9/89 TO 9/93 | 14. DATE OF REPORT (Year, Month, Day)<br>27 Oct. 1993<br>15. PAGE COUNT<br>33 |
| 16. SUPPLEMENTARY NOTATION | |

| 17. COSATI CODES | | | 18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) |
|---|---|---|---|
| FIELD | GROUP | SUB-GROUP | Vision, AI, Cognition, Neurophysiology, Visual Psychophysics, Dynamical Systems |
| | | | |
| | | | |

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

**Although integrated computational and psychophysical studies have considerably advanced our understanding of early visual processing (up to Marr's 2 1/2 D Sketch), much less research, by comparison, is being conducted on intermediate and higher-level vision. One reason for the scarcity is that high-level vision includes goal-directed, context-sensitive, top-down knowledge. However, little is known about these aspects of vision and how they are organised in the visual knowledge bases (for example how default preferences and categorical states are organised and related). Hence plausible, detailed computational models can't be formulated. These studies begin to reveal the structure of some aspects of cognitive visual knowledge.**

| | |
|---|---|
| 20. DISTRIBUTION/AVAILABILITY OF ABSTRACT<br>☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS | 21. ABSTRACT SECURITY CLASSIFICATION<br>Unclassified |
| 22a. NAME OF RESPONSIBLE INDIVIDUAL<br>Dr. John Tangney | 22b TELEPHONE (Include Area Code)<br>(202) 767-5021<br>22c. OFFICE SYMBOL<br>NL |

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted
All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

# Top-Down Influences on Bottom-Up Processing

## 1.0 Introduction

Although much progress was made in the '80's in understanding how surface properties could be recovered from image data, minimal progress was made in recognising objects and events in natural scenes. The exception, of course, was when the domain was well specified and the object classes were known in advance. But without such knowledge, little progress was achieved in obtaining meaningful descriptions of images in terms of objects and their behaviors. Indeed, even indexing to the "correct" object category remained a formidable, largely unsolved task. It became clear that a priori knowledge, to be applied "top down" to the "bottom up" stream of information processors would be necessary. Over the past three years, this research has been aimed at understanding the structure of this "top-down" knowlege.

The work may be conveniently divided into four different areas, plus a fifth part that is a collection of "spinoffs" from the primary effort. The first major area is an understanding of what image features support robust perceptual categories or model classes that have powerful inductive leverage. Critical to such key features is the notion of properties that have special, yet generic structures in the world. The second major advance was a formal definition of a percept – in other words a definition of that state which offers a meaningful description of an image (as opposed to simply a passive symbolic description). Related to both of these areas is the work of Feldman on Perceptual Categories, which constitute the third major area. Finally, we have developed a new psychophysical technique that allows us to probe the categorical structure and special parameterizations of the perceptual feature space.

## 2.0 What Makes a Good Feature?

Let the world consist of various properties $P$ that are associated with various contexts, $C$. Then $p(P|C)$ denotes the conditional probability of a property, $P$, such as "has 4 corners" in the context $C$, which could be sitting "on a plane", "in this region", etc. Similarly the collection of measurements of a property and their conditional probabilities will be specified by $F$ and $p(F|C)$. Note that $p(P|C)$ and $p(F|C)$ are simply objective facts about the world and are *not* statements about the perceiver's model of the world. Our first task is to place conditions on $F$, $P$ and $C$ that ensure the measurements $F$ constitute a reliable indicator that $P$ occurs in the world.

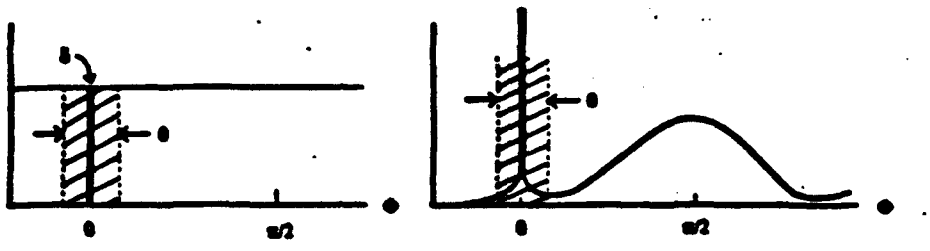



**Figure 1** A: Flat distribution function. B: "Modal" distribution.

## 2.1 Reliable Inferences

The posterior probability of inferring property $P$ given the feature $F$ in context $C$ is $p(P|F\&C)$. A reliable inference makes this probability nearly one, and keeps the probability of an "error", i.e. $p(notP|F\&C)$ near zero. Hence a reliable feature $F$, in context $C$, will keep the following ratio, namely $R_{post}$ much larger than one:

$$R_{post} = p(P|F\&C) \;/\; p(notP|F\&C)\,. \qquad (1)$$

Using Bayes Rule, $R_{post}$ can be broken down into the product of two components, a likelihood ratio $L$ that relates to the "imaging" of $P$ onto $F$, and the prior probability $R_{prior}$, that relates to the genericity of the world property $P$ in context $C$. Specifically, $R_{post} = L \cdot R_{prior}$, where

$$R_{prior} = p(P|C) \;/\; p(notP|C) \quad \text{and} \quad L = p(F|P\&C) \;/\; p(F|notP\&C). \qquad (2)$$

Note that the likelihood ratio captures the intuition that a feature should arise reliably from a given world property, i.e. $L >> 1$. As will be seen in the next section, however, this condition does not insure a reliable inference, because if $R_{prior}$ becomes too small, then $R_{post}$ can become insignificant even in the presence of a high likelihood ratio. (Also see Jepson & Richards, 1992.)

## 2.2 An Example

Consider a world of line segments on a plane seen under orthographic view. Of interest is the special property "two line segments are parallel". Let the threshold for discriminating the orientation difference between two (adjacent lines) be $\theta$, and let $\delta << \theta$ be the limiting resolution of the process that governs straight and parallel. Now let the collective distribution of the orientation $\phi$ of all line segments be rather flat (Figure 1A). Given this context, we are now presented with two lines that fall within the crosshatched sample for $\phi < \theta$. Hence the two lines appear parallel; should we conclude that these lines indeed arise from a parallel process?

First note that the likelihood ratio, $L$, is very high, because (i) whenever parallel lines occur in the world, they always will appear parallel in the image, and (ii) our chance of error is vanishingly small – when two lines are not parallel, they will not be seen as such except in the rare case when they lie within our limit of resolution $\theta$. Hence $p(F|P\&C) = 1$ and $p(F|notP\&C)$ is, say 0.01 if $\theta$ is 1 part in 100. It appears therefore that we should infer that the lines are indeed parallel in the world. However, given our chosen random world context, such an inference is almost always guaranteed to be *wrong*.

Consider the prior probability ratio $R_{prior}$. Because the prior probability $\delta$ of the parallel process occurring is much less than the resolution limit $\theta$, the area occupied by $\delta$ in Figure 1A is much less than the area set by $\theta$. Thus $R_{prior} = \delta/(1-\delta) << \theta$, and its product with the likelihood $L \cong 1/\theta$ will give an a posteriori probability ratio $R_{post} < 1$. Hence the odds really favor the conclusion "not parallel". (See Jepson & Richards, 1992, and Knill & Kersten, 1991.) for further details and examples.) In order to raise $R_{post}$ to a significant level, we need significant priors, say a $\delta$ in this case such that $\delta/\theta >> 1$. In terms of Figure 1, this is equivalent to requiring that the $\phi$ distribution function for pairs of lines be biased, such as indicated in Figure 1B where the process "parallel" appears as a mode in the probability distribution function.

### 2.3 Two Kinds of Regularities

The important message of the previous example is that "good" features arise from some modal regularity in the distribution function of world properties. However, not all regularities satisfying the likelihood and prior conditions will be useful. For example, the property "two skewed lines" satisfies these two conditions, but clearly this property is not very informative. Hence what we seek are properties that are not just arbitrary configurations, but rather ones that are in some sense special.

To illustrate more clearly the fundamental difference between two skewed lines and two parallel lines, we divide structural regularities into two classes: transverse and non-transverse (Poston & Stewart, 1981). Transverse relations arise when the elements of the model are postioned arbitrarily such as the above two skewed lines; non-transverse arrangements require careful positioning, as implied by the term "non-accidental" of Binford (1981) and Lowe (1985). Unlike the notion of "non-accidental", however, the usage of transverse and non-transverse requires a context. Thus, "two parallel lines" (or planes) in a random stick (or planar) world would be non-transverse, but in the context of a building with windows and doors, etc., the concept "parallel" would become transverse. Within the proper context, non-transverse properties are thus very special. But as we showed earlier, in order to be recoverable from image features, the non-transversality must be an isolated spike in the distribution function as in Figure 1B, with sufficient mass to be "visible". This is what previous researchers meant by "modal" properties (Bobick, 1987; Marr,

1970; Richards & Bobick, 1988). Features that satisfy (2) and which arise from non-transverse regularities provide especially reliable and useful inferences about world properties and are called Key Features (see Jepson & Richards, 1992) for "natural" examples taken from motion and color). Loosely speaking, $F$ will be a Key Feature for property $P$ if $P$ is a generic non-transverse mode in the space of world models, and $F$ occurs in the presence of $P$ but never in its absence. Hence the set of properties that image onto the Key Features are an especially useful set of properties, because they are reliably inferable.

### 2.4 An Example

To illustrate a set of properties that image onto key features in our simplified world of line segments in a plane, assume there are two processes that generate two types of relations between two lines. One is the process "parallel"; the other is a process "coincident", where the lines just touch one another. We take these regularities as generic – i.e. we stipulate that both occur with significantly non-zero probabilities in the given context. First we enumerate those regularities that image to key features. Then in the following section, we will place an ordering on this special set of properties.

The enumeration is equivalent to identifying all the non-transverse configurations between line segments in a plane, given the chosen context. We assume the measurement is the orientation of one line to the other, $\phi$, and the position $x$, $y$ of the end-point of one line with respect to the other. Hence the relative positioning has three degrees of freedom (DOF). Referring to Figure 2, the uninformative, transverse regularity chooses $x$,$y$, and $\phi$ arbitrarily, producing two skewed lines. (Intersection or not was not specified in our model class and an "X" will be treated as equivalent to skew without crossing.) First, with care the end of one line can be placed on the other (or its extension), eliminating one degree of freedom. These configurations are assigned a codimension of one. Next, with still more care, we can place the end of one line exactly on the end of the other, allowing only the angle $\phi$ to vary. This arrangement has a codimension of two.

Similarly, if the lines are parallel, then the orientation is fixed and the cost, or codimension of the arrangement is again one. However parallel and coincident lines, with one end allowed to slide along the other, increase the codimension further to two. Finally, we have the last, most special case of positioning of codimension 3 where the two lines merge into one when placed end to end in a parallel arrangement.

## 3.0 From Features to Categories

Our main point will be that the "interesting" structural regularities in a model class – namely those that satisfy the key feature conditions – can be used as a basis for partitioning the model class into categories. In the previous example, the property

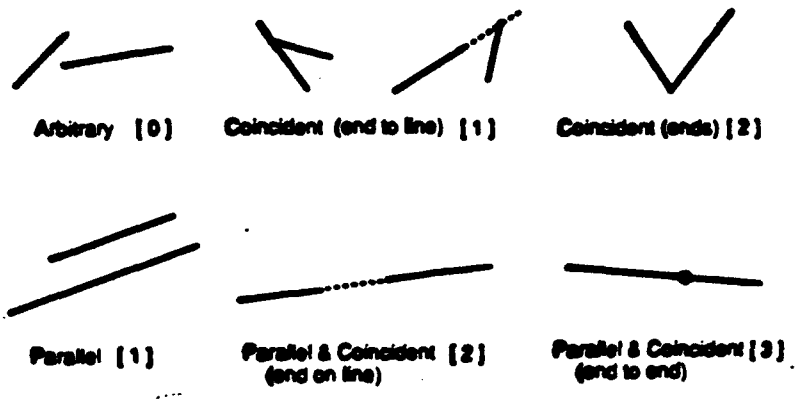



**Figure 2 Line-to-line non-transversalities.**

space would be built from the end-point position measurements $x, y$, and the relative pose $\phi$. Within this $x, y, \phi$ space, our proposal is that the partioning should make explicit the line-to-line non-transversalities illustrated in Figure 2. If this scheme is adopted, then the subspaces will preserve the character of the nontransversal modes, thus distinguishing among the interesting properties. Note that the context sensitivity is critical to our set-up, because it permits legitimate reconfigurations of the property space depending upon the observer's goals, etc.

### 3.1 "Two Stick" Categories: The "Structure Lattice"

Continuing our example, we identify the modal subspace as that associated with our "two-stick" model class presented earlier in Figure 2. Each of these configurations has a codimension, which allows us to place each of these non-transverse modes in a lattice, where each node depicts a proper subspace in the particular context (Figure 3A). The top node shows the arbitrary two-stick configuration. As we move down the lattice, the nodes below differ at each successive level by the removal of exactly one degree-of-freedom from the configuration. Upward transitions, then, are the elemental ones that locally "break" or "unfold" a non-transversal property but which do not add any additional non-transverse properties. An important example is the missing link between the "V" and "parallel line" nodes (or similarly, the "T" and "collinear" nodes). There is no direct route from one node to the other. The explanation is that the concepts "coincident" impose a constraint on the endpoint position $x, y$ of one line with respect to the other, whereas the concept "parallel" is expressed by an angular relation, $\phi$ between the two lines. Because position $(x, y)$ is not defined by angle $(\phi)$ or vice versa in this context, there is no intersection other than the excluded degenerate case of two coincident lines. A similar explanation applies to the missing path from the two "collinear" lines and the "T" node.

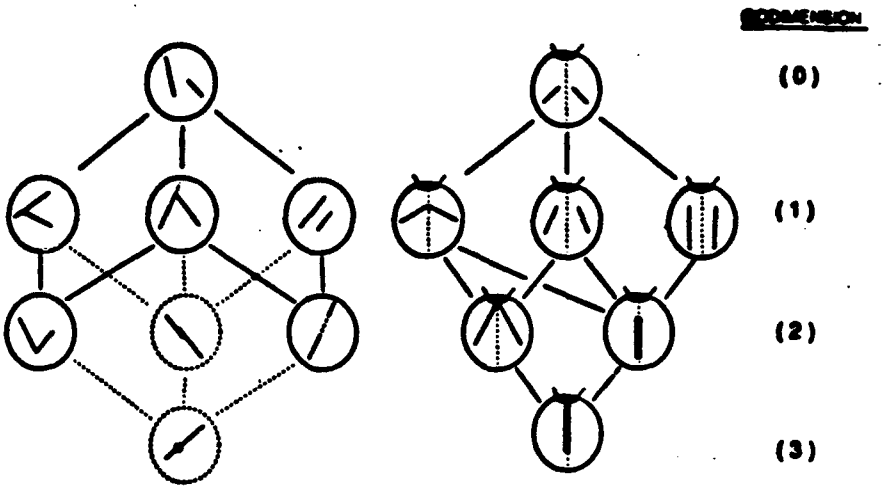



**Figure 3** A: "Dot-on-line" categories. "Two-stick" categories given concepts coincidental andf parallel. See text for explanation of dashed paths and nodes. B: Beetle taxonomy, based on a version of a "two-stick" mode lattice.

At the bottom of the lattice, two nodes have the two sticks collapsed to one. These two nodes have broken outlines to indicate that they are not part of the lattice for the perceptual context because they suggest a "one-stick" configuration. (If the two sticks were each identified in some manner, say by coloring, then the dashed paths and nodes would become part of this "two-stick" category lattice.) Thus, the result of this construction is a lattice that displays a partial ordering of the categorical states available to the perceiver, given the concepts "parallel" and "coincident". We call such an ordering a "structure lattice".

### 3.2 "Natural Example": Beetle Lattice

In the biological realm, growth processes exhibit regularities (Thompson, 1952). To illustrate how such regularities can be used for a taxonomic classification, we will use a simple modification of the "two-stick" mode or structure lattice of Figure 3A. Let the context be the backs of beetle-like bugs that are marked by two distinctive lines oriented with respect to the symmetry axis of the beetle. As is typical for biological shapes, we assume the markings are generated symmetrically about this axis. Hence, with respect to our "two-stick" mode lattice, one stick – the "reference

stick" – will simply be the symmetrical bisector of the beetle's back. The other, namely the "second stick", will thus appear twice in mirror image positions about this symmetric axis. (The situation is equivalent to symmetric markings appearing on a left and right wing.) As before, we assume two possible marking processes, one laying the mark down parallel to the bisecting axis, the other positioning the end point (of either the reference stick axis or the additional marking stick) to be coincident with one of the two lines. All of this sets the context.

Because the two-stick modes in a similar context have already been enumerated, we simply need to recast the previous lattice of Figure 3A in a symmetric form compatible with this revised "biological" context. This has been done in Figure 3B, where now each node depicts the markings on the beetle's back. At the top, the two symmetric marking lines are set arbitrarily with respect to the bisecting axis (dotted). This is the codimension 0 case for this species. At the next level either the coincident or parallel process applies, giving us three codimension 1 subspecies. Next, we have two codimension 2 cases: in one the marking lines form a V, coming together at the "head" of the reference line, or the other where the two marking lines collapse onto the reference line (but do not reach the head of the beetle). Finally we have a single codimension 3 case in this context where the "V" collapses onto the reference bisecting line. Given these generating processes and this context, these are all the types of beetles expected. These types, with the exception of the "generic" beetle at the top, represent the beetle modes or subspecies, each exhibiting a slightly different, but related regularization of the ontology of beetles. Thus the beetle lattice is a convenient hypothesis generator for an observer who is seeking to assign any particular beetle to its "natural" category (Feldman, 1992; Leyton, 1984). Elsewhere, we consider how an observer can induce new categories from the regularized partitioning (Feldman, 1992).

## 4.0 Percepts and Categories

Our basic idea, then, is that the structure and parameterizations of our models describing the world should match the regularities of the image structure as closely as possible. If we wish to extract world structure from image structure in a "vivid" manner (Levesque, 1978), then the properties we especially note in the image should directly point to very specific world properties. This criteria clearly places very strong constraints upon the kinds of image structures we should note, because not all properties of an object can be expected to appear reliably in an arbitrary view of that object.

To clarify this point further, consider the pillboxes shown in Figure 4. In the upper left, most immediately see the handle as rectangular and erect, with the feet lying on the top of the pillbox. However, an infinity of interpretations are possible for this single snapshot. For example, the handle really could be skewed and lying flat on the top – or any other state between flat and erect, including some slants

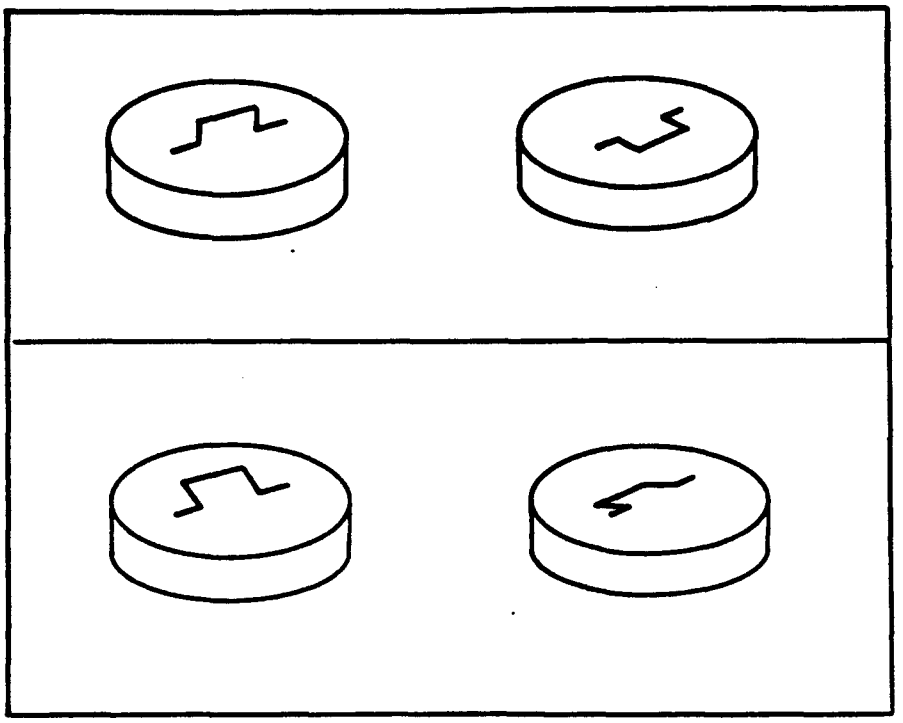


**Figure 4** Some pillboxes with handles. In the upper left depiction, most immediately see the handle as rectangular and erect, whereas in the upper right the handle now appears flat. In the two lower panels, both the shape and inclination of the handle are less clear, the percepts exhibiting some multistabilities. Most favor an inclined rectangular handle for the lower left; the lower right drawing, however, yields mixed reports.

toward the viewer. Nevertheless, observers quickly accept just one interpretation from the many. As will be shown below, such a vivid and compelling perceptual conclusion follows if perceptual categories are built around modal key features.

Let us first examine the orthographic projections of two rectangular handles onto the image plane as illustrated in Figure 5. The normal to a surface N and the visual ray to a point on the surface define a plane perpendicular to the surface at that point. This plane also defines a line in the image. Then the surface normal and any other vector in this plane must project into this image line. The bisector B of a rectangular handle perpendicular to the surface is one such vector. We will define such a handle as an erect, rectangular handle. However, if the same rectangular handle is not erect, i.e. is inclined at some angle to this perpendicular plane, then the angle of its projection is less constrained. In particular, the bisector of a flat handle lying in the plane of the surface can project into *any* angle in the image (see Figure 6). In a random world, where both angles and orientations are cast out with equal probability, the image distribution has a broad spectrum (Witkin, 1981). Clearly, if we had to apply these data to infer the handle shape (i.e. its "skew") and

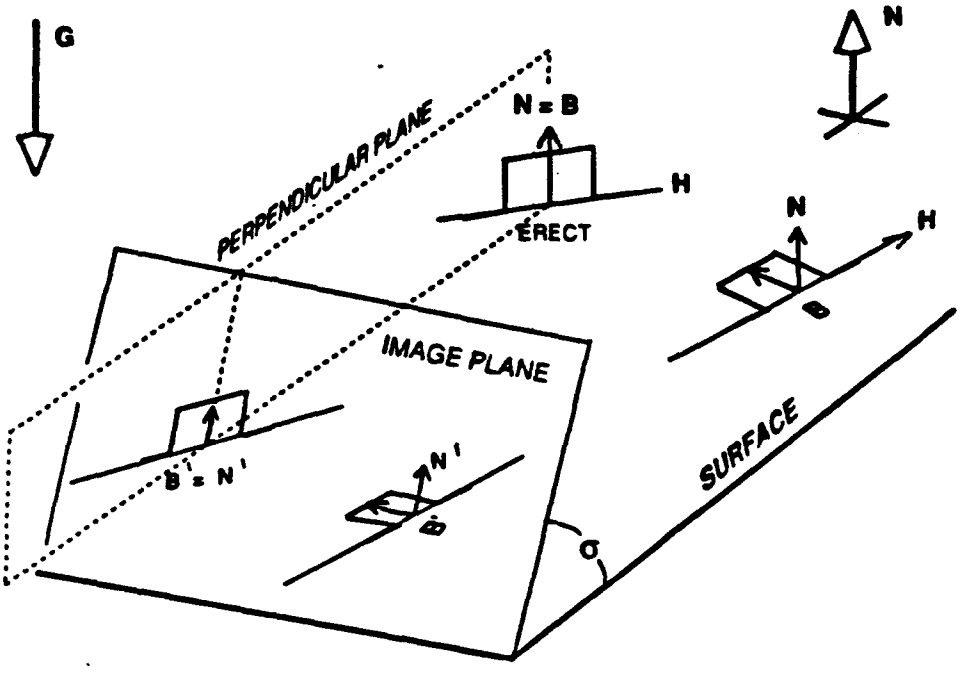



**Figure 5** Regardless of the viewing angle $\sigma$, an erect rectangular handle will project onto the image plane with its bisector B oriented parallel to the projection of the surface normal, N (orthographic projection is assumed). However, if the handle is inclined to the surface or lies flat, then the orientation between the bisector and normal can vary over a wide range, depending on $\sigma$ (see Figure 6).

its attachment angle, at best we could only make a maximum likelihood judgement that would typically be wrong.[1] In order for the perceiver to develop the inferential leverage needed to strongly disambiguate among many possible configurations of equal likelihood, the world must behave somewhat more regularly (Lowe, 1985; Witkin & Tenenbaum, 1985). In particular, some structures should tend to occur significantly more often than predicted by a uniform distribution over all possible structures.

Consider then a world in which the perceiver *knows* that handles will often be rectangular and will lie either flat, as if freely hinged and resting stably under gravity, or erect, as if firmly attached perpendicular to the surface. In this world, the distribution of handle orientations $\alpha$, rather than being uniform, will have two "spikes" or "modes", one at each of the two special world configurations as shown in Figure 6A. Now, depending upon the slant of the surface, the expected image distribution of the handle bisector will be as in Figure 6B: the "erect" bisector

[1] Surprisingly, given no other information, the maximum likelihood estimate for the 3D angle is just the image angle itself!

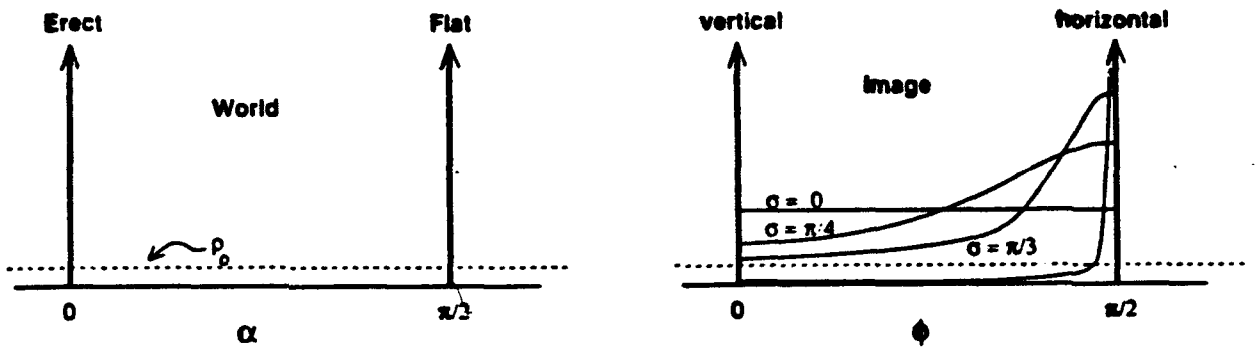



**Figure 6** **Left: Two states of a rectangular handle are taken as regularities in the world: an erect state where the plane of the handle is perpendicular to the top surface of the pillbox and a flat state where the plane of the handle coincides with this top surface. The angle $\alpha$ is the angle between the bisector of the handle B and the surface normal N. The dotted line labelled $\rho_o$ indicates the density function for arbitrary angles, other than the erect (0) and flat ($\pi/2$) regularities which have spikes in the probability density function. In the image (right), the erect handle also has a spike in the density function for orientation, because parallel vectors in the world are parallel in the image, hence the image angle $\phi$ of B′ to N′ is 0. All other handle inclinations project onto image angles that depend upon the viewpoint, or "slant" of the surface ($\sigma$).**

continues to stand out distinctively. As discussed in the previous section, such an image feature is designated a "key" feature because (i) its likelihood of correctly indicating the presence of a particular world property is high (i.e. there are few false targets), and (ii) the associated world configuration has a significant prior probability (see Knill & Kersten, 1991; Jepson & Richards, 1992). This latter condition, though often overlooked, is critical to establishing that a given high-likelihood world interpretation is actually likely to be correct. In other words, the configuration ascribed to the world by an inference must actually be one that commonly occurs in the context. Otherwise, the probability of the inference being correct will actually be dominated by the probability of a false target.

### 4.1 Structure Lattice

To set up our representation, we begin by introducing the vehicle called a "structure lattice" that takes our context-sensitive, primitive concepts about structural regularities, and composes them to produce a set of possible configuration states. (Just as we did earlier for the "beetles".) This is the first of several such lattices we will introduce, the one upon which the later lattices will be built. Each of these lattices will display a partial ordering of the categorical states. (See Moray, 1990, for a related proposal.) In the case of the structure lattice, the ordering is derived

by noting that some states are special or limiting cases of others. Later we will impose context-specific preferences upon this collection in order to seek a maximally preferred state.

To illustrate again in detail the role of regularities in creating a representation in which the perceptual categories become obvious, let us propose that alignment and perpendicularity be special regularities between lines (or vectors) that we encounter in our non-random world. For example, assume that object parts have coordinate frames that are often aligned in some manner (Arnold & Binford, 1980). For the pillbox and handle, we have two "parts" and hence two coordinate frames. Let us specify the coordinate frame for the pillbox by its symmetry axis **A**, and by the feet of the handle **H**. (See Figure 7.) We will assume that the pillbox has been cut at right angles to **A**, and hence the surface normal **N** to the top of the pill box will align with the axis **A**. (Note this assumed axiomatic regularity!) Together, **A** and **H** (or henceforth **N** and **H**) set up a right-angled Cartesian coordinate frame at the center of the top of the pillbox. Let **K** be a unit vector orthogonal to **N** and **H**, defined by $\mathbf{K} = \mathbf{N} \times \mathbf{H}$. To construct $\mathbf{K}'$, the projection of **K** into the image, we use the maximum likelihood rule for slant derived by Kanade (1983), which was observed psychophysically by Stevens (1983) for right-angled coordinate frames. Specifically $\mathbf{K}'$ will lie on the bisector of the angle between **N** and **H**, illustrated in Figure 7. (Shortly we will explain the role of the numbered sectors marked on the top of the pill box.)

Similarly, a coordinate frame for the handle can be defined by its vertical symmetric bisector **B** and by a second vector **H** which is the direction of the feet of the handle. Note that we do not assume that **B** and **H** are perpendicular. However the origins of the two coordinate frames, **B** & **H** and **N** & **H**, are assumed to lie centered on the plane of the top of the pillbox, and coincident with the major axis of the pillbox. We thus are assuming the following:

## Contextual Regularities:

| | |
|---|---|
| Parts: | Pillbox is convex (e.g. solid top).<br>Handle is planar. |
| Support: | Both feet of handle lie in plane of top surface of of pillbox (**B** lies on or above this plane). |
| Surface Normal Alignment: | $\mathbf{N} = \mathbf{A}$ |
| Gravity Alignment: | $\mathbf{A} = \mathbf{G}$ |
| Cartesian Frame: | $\mathbf{N} \cdot \mathbf{H} = 0$ |
| | $\mathbf{K} \cdot \mathbf{H} = 0$ |
| Viewpoint: | Pillbox is seen from above. |

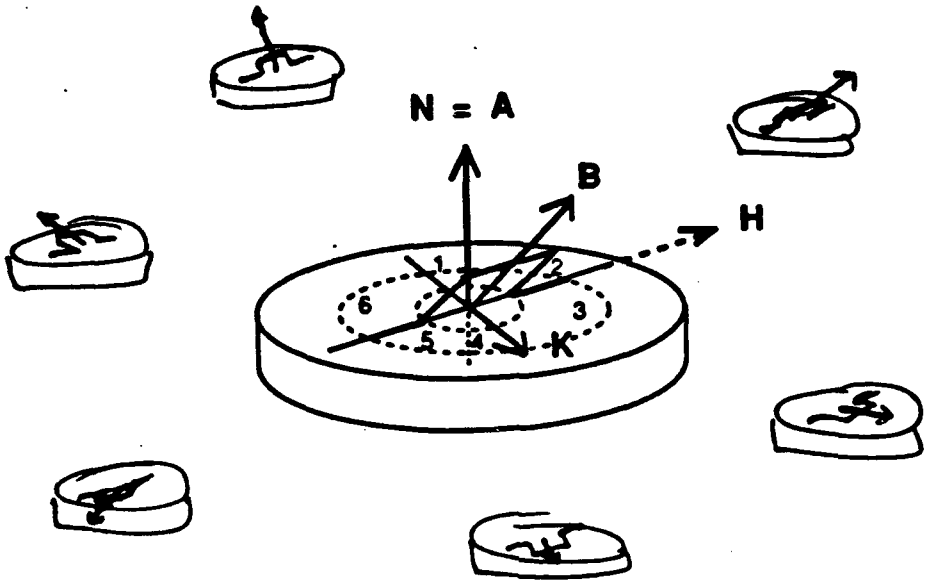



**Figure 7** Unit vectors that define the handle and pillbox coordinate frames. The diagonal line that bisects N, H sets up the third right angled axis K. The dotted ellipses are circles that create the six sectors discussed in the text. The insets depict handles with bisectors projecting into the various sectors.

The additional vector G is taken to be the gravity axis, which is aligned with the customary page orientation, typical for the depiction of a stably supported object.[2] In sum, the above equalities set up two coordinate frames, one rectangular for the pillbox defined by N and H and the other not necessarily rectangular for the handle defined by B and H.

Given the vectors N, B, H and K we can now explore all possible alignments of these vectors. Recall we are proposing that the perceiver is aware of certain "modes" or configurations of structures that occur often in the world. In particular the special regularities we chose were the collinearity of two lines or vectors, such as $\mathbf{B} = \mathbf{N}$, and the perpendicularity of two lines, such as $\mathbf{B} \perp \mathbf{H}$, which corresponds to a rectangular handle, or $\mathbf{B} \perp \mathbf{N}$ which defines a flat handle. Hence to generate all these special configurations that are the consequence of these particular relational concepts, we simply enumerate all the alignments of B with N, K and H, using either the collinear (=) or perpendicular ($\perp$) relation. The result of this enumeration will then be those special categories that make sense to us, given our chosen relational concepts. We begin first with the three collinear alignments:

[2] Elsewhere we have explored this preference for supported objects (Jepson & Richards, 1993).

| Collinear Relation | Category | Notation |
|---|---|---|
| $\mathbf{B} = \mathbf{N}$ | erect rectangular handle | $ER$ |
| $\mathbf{B} = \mathbf{K}$ | flat rectangular handle | $FR$ |
| $\mathbf{B} = \mathbf{H}$ | degenerate (infinitely skewed handle) | |

If the bisector **B** does not align with either **N**, **K** or **H**, then we define the handle as being either "tilted", which is noted as "$T$", or "skewed", which is noted as "$S$", or both, namely "$TS$". In particular, if the bisector is in the plane determined by **N** and **K**, then the handle is tilted and rectangular, i.e. "$TR$". Similarly, if the bisector is in the plane containing **N** and **H**, it will be erect and skewed, i.e. "$ES$", while for the flat and skewed state the bisector will be in the **H**-**K** plane. Thus, excluding the above collinear specializations, we now have the following additional three new cases (alternatively we could have filled out a $4 \times 4$ table):

| Perpendicular Relation | Description | Notation |
|---|---|---|
| $\mathbf{B} \perp \mathbf{H}$ | tilted rectangular handle | $TR$ |
| $\mathbf{B} \perp \mathbf{K}$ | erect "skewed" handle | $ES$ |
| $\mathbf{B} \perp \mathbf{N}$ | flat "skewed" handle | $FS$ |

Finally, we have the category where none of the relations hold:

| Arbitrary Relation | Category | Notation |
|---|---|---|
| (none of the above) | tilted, skewed handle | $TS$ |

Excluding the degenerate case $\mathbf{B} = \mathbf{H}$, we thus have six types of categories for the positioning of the handle, given our conceptualization that part-based structures in the world typically are related by an alignment of some aspect of their individual coordinate frames. Because we can count the number of axes of each frame that are aligned (i.e. either one axis or two), a partial ordering can be placed on these six types of structures. This is illustrated in Figure 8 as a graph or lattice. At the top of this lattice, the positioning, $T$, and shape, $S$, of the handle is arbitrary. At the bottom, however, we have two states where the position and shape of the handle are both fixed to be rectangular and either flat or erect (i.e. $FR$ and $ER$). In other words, all degrees of freedom of alignments have been removed. In between are the planar alignment states where one degree of freedom of movement is still allowed. For example, the leftmost node $ES$ permits the skew of the handle to vary, but it must remain erect. Hence, as we move from top to bottom in this lattice, more and more specialization or restrictions are placed on the configuration. As previously mentioned, we call this lattice a "structure lattice" because, given the assumed alignment regularity this lattice shows the specialization relations between the categories of structures that will appear in our representation.

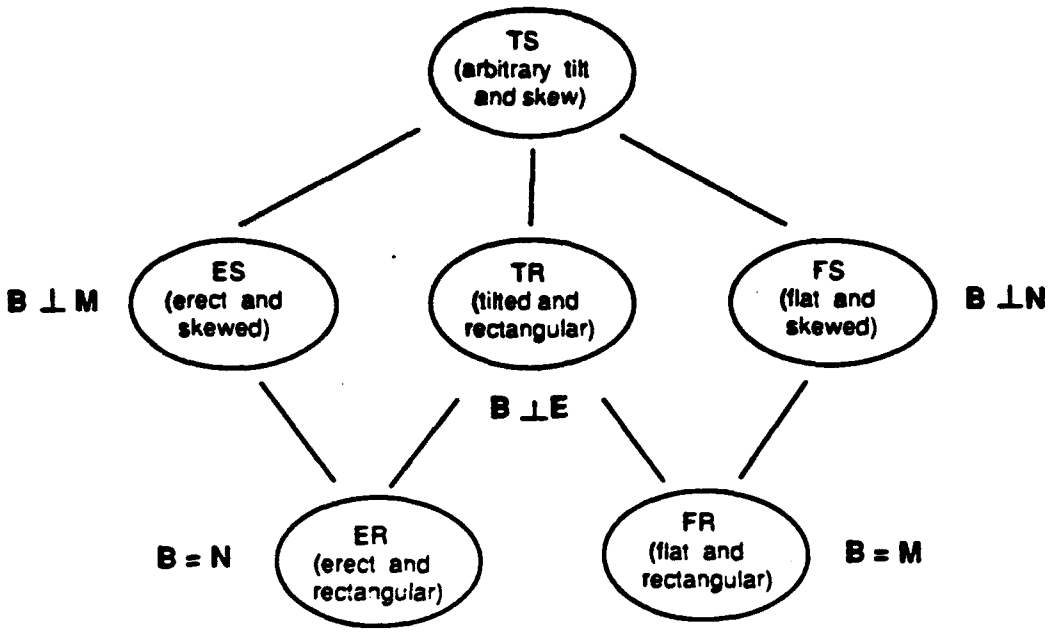



**Figure 8** The structure lattice for the pillbox plus handle (i.e. the "state space").

## 4.2 Preference Relations

The structure lattice simply enumerates the structural categories that we know about, or can easily infer, given our chosen regularities. Ideally, we would hope that the image is consistent with some kind of maximization of these regularities. In other words, given a particular context, we expect certain regularities to appear, but in another context the structures expected might differ. For example, a "flat" handle would not be likely if the pillbox were upside down. This suggests that given a context, there is a preference ordering on the expected regularities. If you will, a ranking is given to the prior probabilities of the structures that are expected in the assumed context.

A preference ordering differs from the structure ordering introduced in the previous section. The structure lattice simply presents all the categories available to us in the chosen context, ordered with respect to increasing specialization of structure. A preference ordering specifies which kinds of specializations are preferred to others. So, for example, given a choice between handle shapes that are rectangular or skewed, we'll prefer the rectangular version. This preference is not surprising, because, after all, we elected to parameterize the coordinate frame for handle shape in terms of rectilinear coordinates. Denote this preference for rectangular over skewed shapes as $R > S$. Similarly, for the attachment angles, our parameterization suggests that the erect "E" and flat "F" angles will be preferred

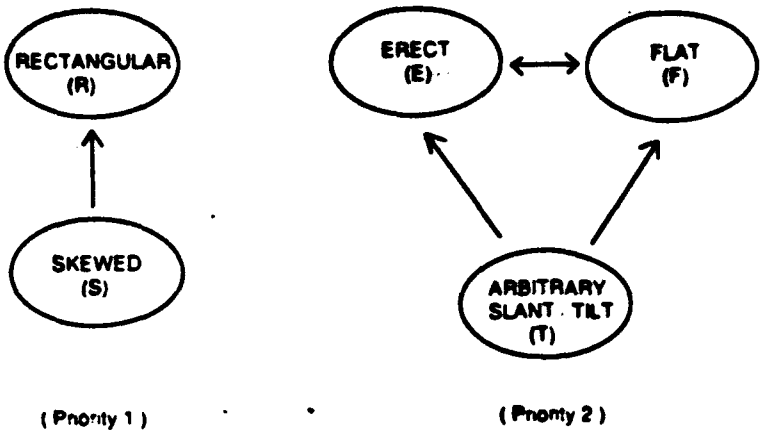



**Figure 9 Elemental preference relations for handle shape (left) and handle inclination (right).**

over arbitrary inclinations, or "tilts", "T", hence $E > T$ and $F > T$. However, we have no reason to believe that an *arbitrarily* shaped erect handle "E" will be preferred over one that is flat, "F". Denote this indifference by $E \sim F$. We will designate these three orderings, one for shape and the other two for attachment angle as the "elemental" preference orderings. They can also be cast in the form of a directed graph as in Figure 9.

Using the above elemental preference relations, we can now impose a partial order on the states of the base plus handle configurations that we know about, namely the states shown in Figure 8. This preference ordering is based on the *consensus* of the elemental preference relations, and is illustrated in Figure 10. Note that a state such as $ER$ is to be preferred over $TS$ because both of the elemental preference relations, $E > T$ and $R > S$, favor the same state. However, such a consensus does not always occur. For example, the same two elemental relations are in conflict for the states $ES$ and $TR$, and as a result these two states remain unordered in the preference ordering. The intuition behind such unordered states is that the perceiver does not have sufficient information to be able to resolve whether $ES$ should be preferred to $TR$, or vice versa. Thus unordered states represent a total lack of information on the appropriate preference. In addition, we also have a distinct notion of an equal preference between two states, such as occurs between $ER$ and $FR$, as well as between $ES$ and $FS$.

In general we cannot expect a consensus ordering to provide a total ordering of the state space, because some conflicts amoungst the elemental preference relations are likely to hold. This is related to Arrow's general impossibility theorem which states that rational choice - i.e. rational voting behaviour - does not guarantee a unique winner (Doyle & Welman, 1989; Saari, 1992). Somewhat counter-intuitively, the introduction of more elemental preference relations does not lead to a more

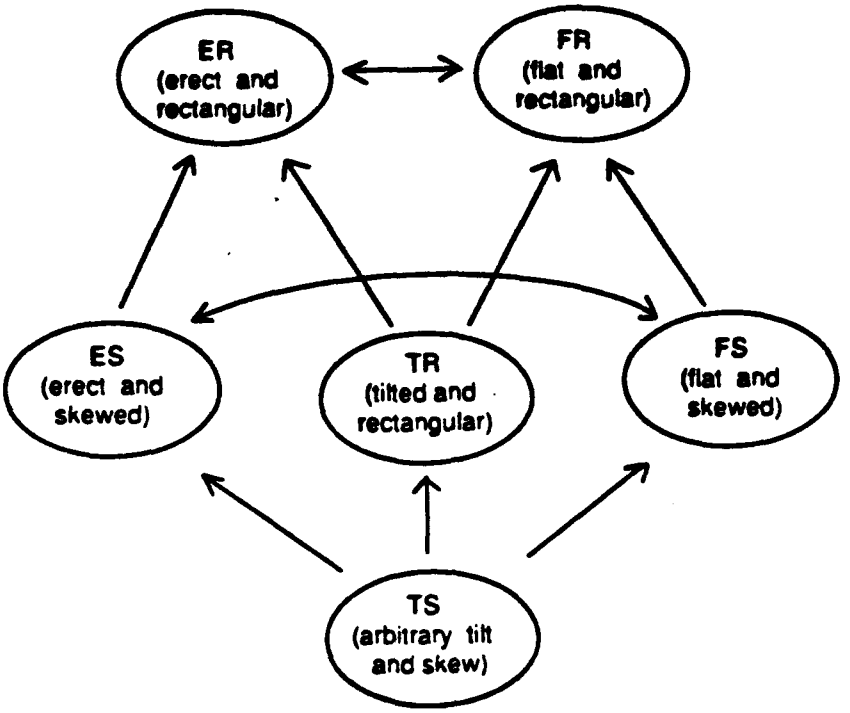



**Figure 10 A reordering of the entire "state-space" based on the assumed elemental preference relations.**

complete ordering, but rather tends to introduce more conflicts and hence tends to eliminate ordering relations. To counter-act this tendency to fracture the state-space, it is often useful to consider priorities amoungst the elemental preference relations (Jepson & Richards, 1993). Such priorities can break particular conflicts and thereby enlarge the ordering. Nevertheless, we should expect typical preference orderings to be partial as a consequence of the incomplete knowledge a perceiver has of its current domain. Of particular interest are instances in which the ordering results in several *maximally* preferred explanations of the image structure, where it remains undecided just which maximal state is to be preferred. As we discuss elsewhere, this is an intuitive explanation behind the difference in the stability of the percepts in the upper and lower panels of Figure 4.

### 4.3 The Pillbox plus Handle

To clarify our framework further, we now return to Figure 4, and use these images together with the preference relations to impose an ordering on the state space in each case. For all depictions, we assume that the view is from above. We also initially assume that a Cartesian coordinate frame for the pillbox is set up using the Kanade-Stevens rule, namely that the axis K is seen as lying along the bisector

of the image angle between N and H (i.e. as shown in Figure 7). We take this coordinate frame as being the *unique* coordinate system containing the line H and the line perpendicular to N. Later we will consider cases when this frame itself appears as a preference that may be altered.

We begin by choosing a representation that allows us to effortlessly read off the states of the handle of interest, given a particular image. Figure 7 shows the form of this vivid representation, based on the N, K and H coordinates. The added feature is that now we identify the six sectors of unit circle (seen slanted) that are seen to lie between the projections of these axes of the coordinate frame. The idea then is to regard the bisector B as the arm of a clock, and simply note either the sector it falls in, or whether it is precisely aligned with one of the axes. A simple example is when B is aligned with either N or K. If B = N, then obviously the erect rectangluar handle *ER* is a possibility, because the handle is *ER* if and only if B = N, whereas if the handle is flat and rectangular, then B = K. Similarly, if B falls into one of the six sectors, again we can easily check to see if a state is consistent or not. For example, when the handle is rectangular and tilted forward, B must be in the upper quadrant of the NK plane and hence its projection must fall into sectors 2 or 3 (see insets to Figure 7). Similarly if the handle is erect but skewed, B must lie in sectors 1 and 6 (if skewed to the left) or sector 2 (if skewed to the right). The following table captures all the cases (excluding the alignments):

| Sector | Possible Categories |
|---|---|
| 1 | *TR* (backward), *ES*, *FS*, *TS* |
| 2 | *TR* (forward), *ES*, *FS*, *TS* |
| 3 | *TR* (forward), *FS*, *TS* |
| 4 | *FS*, *TS* |
| 5 | *FS*, *TS* |
| 6 | *ES*, *FS*, *TS* |

**Table 1** The possible attachment categories for handle pose, given the sector that the bisector B falls into. (See Figure 7.)

Note that our condition that the handle lies on or above the top of the pillbox constrains *TR* and *ES* to require that B not fall in sectors 4 and 5.

The state space for the two upper drawings in the top panel of Figure 1 is given in Table 2. Again, we use the notation *E*, *F*, *R* respectively to indicate an erect, flat or rectangular handle, or *S* and *T* repectively to indicate either a skewed or "tilted" handle. First consider the possibilities for the "erect" handle depiction in the upper left drawing. The bisector B aligns with N. Hence *ER* is an obvious choice. However, B can also lie off N, but in the plane defined by the visual ray.

| Upper Left Drawing | Handle State | Upper Right Drawing |
|---|---|---|
| TS | arbitrary tilt & skew | TS |
| * | erect, skewed handle | * |
| FS | flat, skewed handle | (FS) |
| * | tilted, rectangular | (TR) |
| * | flat, rectangular | FR |
| ER | erect, rectangular | * |

**Table 2** State spaces for the two drawings in the upper panel of Figure 4.

All of these states correspond to either tilted and skewed ($TS$) handles, or perhaps a flat and skewed ($FS$) handle. Note that erect and skewed ($ES$) is not consistent with the Kanade-Stevens coordinate frame since, from Figure 7, we see the only way the handle can be in the NH plane (i.e. erect) yet have B align with N in the image is for B to equal N (i.e. the $ER$ state). Similar arguments showing that $TR$ and $FR$ states are inconsistent can also be read off of Figure 7. These three inconsistent states are indicated by an asterisk in Table 2. The remaining three valid states, $TS$, $FS$, and $ER$ can now be ordered using consensus amoungst the elemental preference relations introduced above. The result is shown in Figure 11 left, and is seen to be a total ordering with the erect rectangular handle ($ER$) as the unique maximal state.

Similarly, for the upper right drawing we first note that the leg of the handle, hence the bisector B appears to align with the axis K in the Kanade-Stevens coordinate frame for the pillbox. Therefore, $FR$ is obviously in the state space. However, the true 3D orientation of B need not be coincident with K, but can point anywhere in the plane created by the lines of sight through K, and hence $TS$ is also a possibility. Obviously the erect states $ES$ and $ER$ are excluded because B lies in sectors 3 and 4 *below* H. States $TR$ and $FS$ are marginal, depending on whether B is taken to be precisely aligned with K or not. If B is seen to fall below K in the representation depicted in Figure 7 (i.e. in sector 4), then $TR$ is not in the state space. But if B lies above K (in sector 3), then $TR$ is a possibility. In either case, $FS$ is possible. Because of this ambiguity $TR$ and $FS$ are shown parenthetically in Table 1, and as dotted nodes in the preference ordering of Figure 8 (right). Again, the ordering here follows from the relations $F > T$ and $R > S$, yielding the state $FR$ seen most "vividly" as the maximal node.

For the two more ambiguous drawings in the lower panel of Figure 4 we may go through a similar exercise. This is treated elsewhere (Richards & Jepson, 1994).

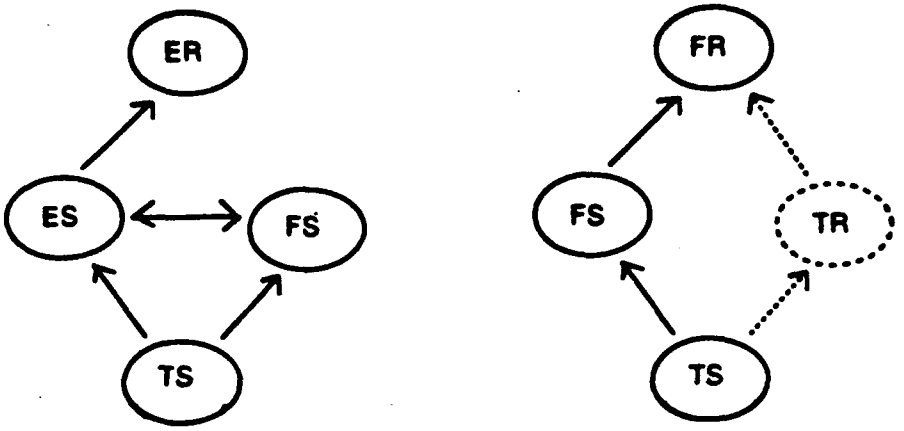



**Figure 11** A preference ordering for the two drawings in the upper panel of Figure 4. The two dotted nodes at the right become possibilities if B is not precisely aligned with K.

### 4.4 Summary

Our notion then is that each image is evaluated with respect to the current set of observed regularities. These regularities suggest a context that dictates the form of the model representation. Given this representation and the image, a set of categorical structures can be deduced easily as "vivid" states (i.e. the state space). The context also points to preferences for certain 3D regularities in the representation, which are used to place an ordering on the feasible states or "interpretations". Hopefully there will be a unique global maximum in this ordering that "explains" all the observed regularities, given the image and the preferences (such as in Figure 11). If not, or if further regularities are observed in the resultant 3D interpretation, or if additional relevant premises are retrieved from the knowledge base, the context may be revised and the process continued with the aim of insuring that all regularities, both in the image and in the interpretation, are explained by the preferences at hand. Sometimes, as in the lower panel of Figure 4, closure is not possible, and several maximal interpretations continue to be evaluated. In all cases, the explanation of the image attempts to maximize our preferences for certain world regularities over other states. This leads to the following proposal for defining a "percept":

> **Proposal 1:** given a context, a percept is an interpretation in the state space that is locally maximal within the associated preference ordering.

Elsewhere (Jepson & Richards, 1993), we have elaborated the consequences of this proposal, and its implications for the machinery that underlies the perceptual process itself. However, of special interest may be the relation between the above Boolean proposal for percepts and one based on versions of utility or probability theory (such as Dempster-Shafer or Pearl's Bayesian graphs). A partial bridge to these alternative approaches using a Bayesian formalization is in the press (Richards & Jepson, 1994).

## 5.0 Trajectory Mapping (TM)

According to our view, special features, called "key features", are critical to setting up perceptual categories and to driving our percepts. Once these features are chosen, then the parameterizations in the perceptual space – the feature space if you will – are forced. Over twenty years ago, multidimensional scaling was introduced to recover the parameterizations used by the human perceiver in a variety of domains. Recently, we have invented a new scaling procedure, called "Trajectory Mapping" that specifically follows the feature paths in the perceptual space (Richards & Koenderink, 1993A).

Briefly, TM is based on the estimation of smooth trajectories in a multidimensional feature space. Given a set of samples, pairs are used as fixed points on a trajectory that is indicated by choosing the most appropriate extrapolants (interpolant) from the remaining samples. A distance measure for these choices is estimated using the current two fixed points as the unit. Unlike MDS, estimates are not made if the fixed point samples are deemed incompatable. Also, samples that lie on a bound in the space – either interior or exterior – are also noted. Hence, trajectories are not estimated across mutually exclusive features, such as the "red-green" exclusion in color. The result is a web of smooth trajectories between samples in the feature space – something akin to a subway map – with some of the trajectories terminated either at the boundaries of the space, or at the exclusion boundaries. These trajectories of feature similarities are then scaled in a higher dimensional space that produces a hyper surface on which the trajectories correspond to "least energy" paths. For example, three point samples on a spherical surface with no exclusions could produce four trajectories. Three are great circles and the fourth is the intersection of the sphere with the plane containing the three points. A hyperbolic surface will create different paths in the feature space. To compare MDS and TM, we present scaling results for 20 OSA uniform color samples (*Jrl. Opt. Soc. Am.* **64**, 1691, 1974). A ridge along the yellow-blue axis appears, which both deflects and attracts the trajectories, which do not cross this ridge. The neutral gray point also appears as a post through which no trajectory passes. Similar behaviors are noted when the TM method is used to scale thirty natural textures taken from the Brodatz album (Dover, 1966). A complete report on this method and these preliminary results will appear shortly (Richards & Koenderink, 1993B).

## 6.0 Dynamical Systems Analysis: Is Perception Chaotic?

The multistability of impoverished visual displays, such as the Necker Cube or the crater illusion presented in Figure 12, is well known. Less appreciated is the fact that even complex scenes, although rich in visual information, also may have many interpretations – indeed perhaps even an infinity depending upon the level of detail desired to "explain" the scene. Different patterns of eye movements are often associated with these alternate interpretations, as shown by Yarbus (1967). Hence, despite our impressions to the contrary, the chosen percept typically entails the selection of one out of many possible interpretations of the input, even if the context remains unchanged (Jepson and Richards 1993). Here we present evidence that processes searching for the appropriate percept have some characteristics associated with low dimensional, chaotic dynamical systems. Our results confirm a proposal by Poston and Stewart (1978) and conclusions reached by Ta'eed et al. (1988) that multistable percepts can be modelled as non-linear dynamical systems.

### Method

A dynamical system can be characterized by the geometry of the space of its output. These outputs are typically a sequence of state changes, either temporal or spatial. Hence possible measures that we can use to evaluate a perceptual dynamical system are (1) the sequence of time intervals between perceptual states, such as the duration of successive flips of a Necker Cube, or (2) the sequence of spatial vectors, such as when the eye moves from one $(x, y)$ position to another. In either case the data will be a sequence of values, $\{v_i\}$ for $i$ typically greater than 200. To determine whether an observed pattern exhibited chaotic behavior characteristic of a dynamical system, we chose a correlation technique outlined elsewhere (Bergé et al. 1984; Grassberger and Procaccia 1983). This technique has been especially successful in using time series data to analyze non-linear physical systems, such as turbulant flow (Essex, Lookman & Nerenberg, 1987; Malraison et al. 1983) or semi-conductor resonators (Van Buskirk & Jeffries 1985; Theiler 1990). Briefly, the method first computes the number of vectors of length $(\mathbf{x}_i - \mathbf{x}_j)$ falling within a $p$-dimensional hypersphere of radius $r$, where $\{\mathbf{x}\}$ is the set of $m$ sample points (e.g. time intervals) collected:

$$C_p(r) = \frac{1}{m^2} \Sigma_{i,j=1 \text{ to } m} H[r - |\mathbf{x}_i - \mathbf{x}_j|]$$

$C_p(r)$ is the average number of vectors or the "correlation coefficient" for the chosen radius, $p$ is the "embedding" dimension, $m$ is the total number of values in the sequence and $H$ is the Heaviside function, which is evaluated to one if the distance between the pair $i$, $j$ is less than $r$, otherwise zero. In other words, the correlation function counts the number of pairs with distance $|\mathbf{x}_i - \mathbf{x}_j|$ smaller than $r$, where $r$ is the radius of a $p$-dimensional hypersphere. The method thus provides an assessment of the geometry of the space from which the samples are taken. For example, if the

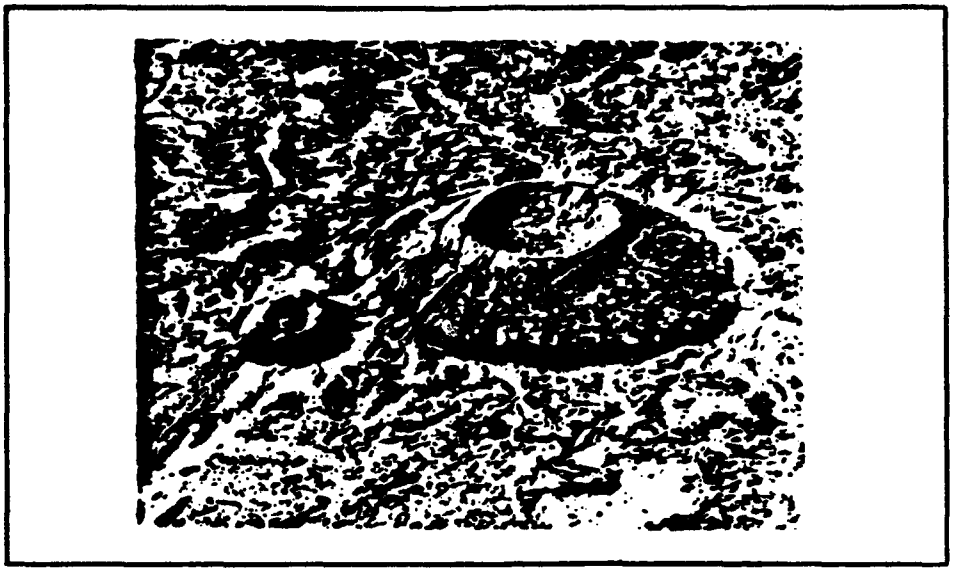


**Figure 12** Although the photograph is actually of two cones in the desolate Ka'u Desert lava field in Hawaii, occasionally the alternate interpretation of two craters is preferred. (This preference can be reinforced by turning the picture upside down.) [Courtesy of Wide World of Photos, Nov. 1972.]

samples are uniformly distributed on a plane, then increasing $r$ will increase the correlation count as the square of the radius. For each embedding dimension $p$, an exponent $e_p(r) = \log[C(r)]/\log[r]$ is calculated over $r$. The maximum value of this exponent $e_p$ is then determined and plotted against $p$. If a random time series is evaluated by this method this exponent rises with the embedding dimension (i.e. $e_p = p$, or more correctly, $e_p \doteq p$ for stochastic series). If a deterministic chaotic series is evaluated, such as that generated by a Henon attractor, then the maximal values of $e_p$ asymptote at some $p_{\max}$ for all $p \geq p_{\max}$. For a typical predator-prey dynamical system, such as May's logistic function, $p_{\max} \doteq 1.2$. Figure 13 illustrates this method. For each $p$, $\log C_p(r)$ was plotted versus $\log r$ and the steepest slope $e_p$ was computed using 6 points for $r$ separated by factors of 1.2 (i.e. $2^{\frac{1}{4}}$). Note that as $p$ increases, so does $e_p$. These slopes were then plotted versus $p$, as shown in Figure 14. The error bars show the range of slopes, based on adjacent points, within the six-point average plotted as a filled circle. The weakness of the method is also illustrated here: for large embedding dimensions, $p > 5$, the uncertainty in $e_p$ increases substantially, especially when the number of data points is small (i.e. $N \doteq 200$). Hence in our study estimates of $e_p$ are restricted to $\leq 8$.

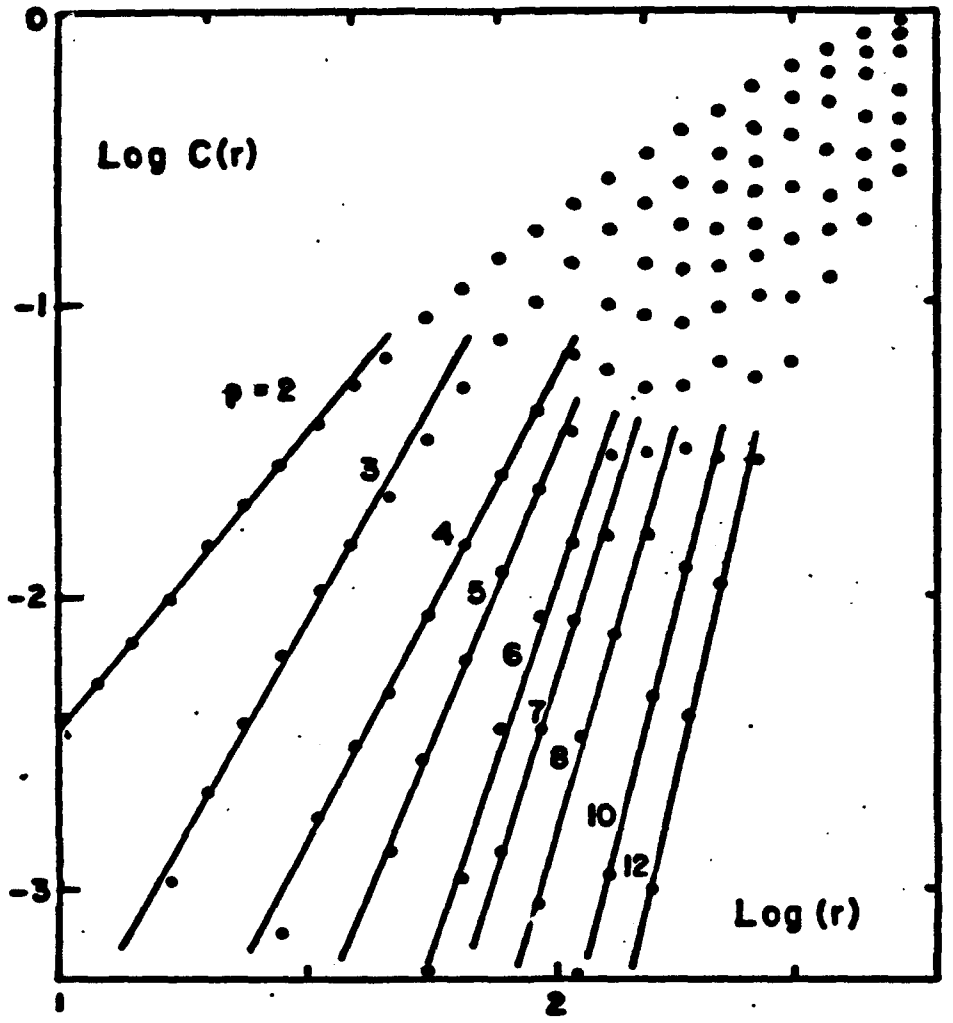



**Figure 13** First stage in the analysis of the time sequence of state changes for the crater illusion of Figure 1. Abscissa is the radius $r$ of the hypersphere of dimension $p$. Each point shows the value for that $C_p(r)$. The slopes on this log-log plot yield the values for $e_p$.

## Experiments

We obtained data for four sets of perceptual tasks: binocular rivalry; fixation patterns during search; simple multistable percepts; and perceptual segments or "eras" in several movies. The panels in Figure 15, plus Figure 12, show examples of the multistable stimuli used. Table 1 gives the number of reversals recorded, and the mean reversal time.

## Binocular Rivalry

When two patterns having significantly different spatial structure are presented, each to separate eyes, typically one sees an alternation of one pattern with the other. For example, if one eye is presented with vertical red and black bars, whereas the other sees horizontal green and black bars, the percept is dynamic alternation between each. The site of this competition or "rivalry" is known to be early in the visual pathway, namely near the initial cortical area of processing (Julesz 1971; Richards 1970). We measured the time series for 200 alternations for two subjects and computed the correlation coefficients and exponents $e_p$ for embedding dimensions $p = 1$ to 8. As shown in Figure 14, a plot of $e_p$ vs $p$ is roughly a line of unit slope, specifically 0.85 for this subject. This result, regardless of the underlying cause, is in agreement with Levelt (1965) and is typical of a stochastic process such as the pattern of rainfall. Hence we conclude that binocular rivalry is not a deterministic chaotic process typical of a predator-prey competition (Kadanov 1986; May 1976).

## Cognitive Rivalry

Unlike "passive" binocular rivalry, the flip-flops of the crater picture in Figure 12 reflect processes that are evaluating the image as a depiction of a world state or "scene". The competition, if you will, is between alternate models that "explain" the image. It is easy to imagine a process, such as the fluctuations in predator-prey populations, where one explanation dominates for a while, and then another takes over. We studied several simple displays such as the illusion of Figure 12, a picture of a triangle with a stick lying across one edge, as well as some cognitive rivalry tasks invented by Manfred Fahle (Fahle and Palm 1990). All of these examples revealed hints of chaos. One of the clearer results was obtained from the time series of flip-flops for the crater illusion. This display is particularly well-suited for study because the perceptual biases can be manipulated by rotating the pattern, placing the illuminant above/below or left/right – the latter being chosen because it yields a more equal competition between percepts.

The data for observer AJ are presented in Figure 14, upper right. (The results for WR were similar – see Table 3.) Note that, in contrast to binocular rivalry, we do not see $e_p$ increasing linearly with $p$, but rather rising and leveling off near $e_p = 3$ for $p \doteq 4$, then rising again. This "notch" in the curve is significantly different from the linear extrapolation of the points for $p > 5$. We propose that this "notch" results from a process exhibiting deterministic chaos of dimensionality roughly 3.5, corrupted by a noisy stochastic process that "takes over" as the embedding dimension increased.

Note that data for the reversals of a Necker Cube do not appear in Table 3. As was found for binocular rivalry, only the hint of a "notch" was observed in the

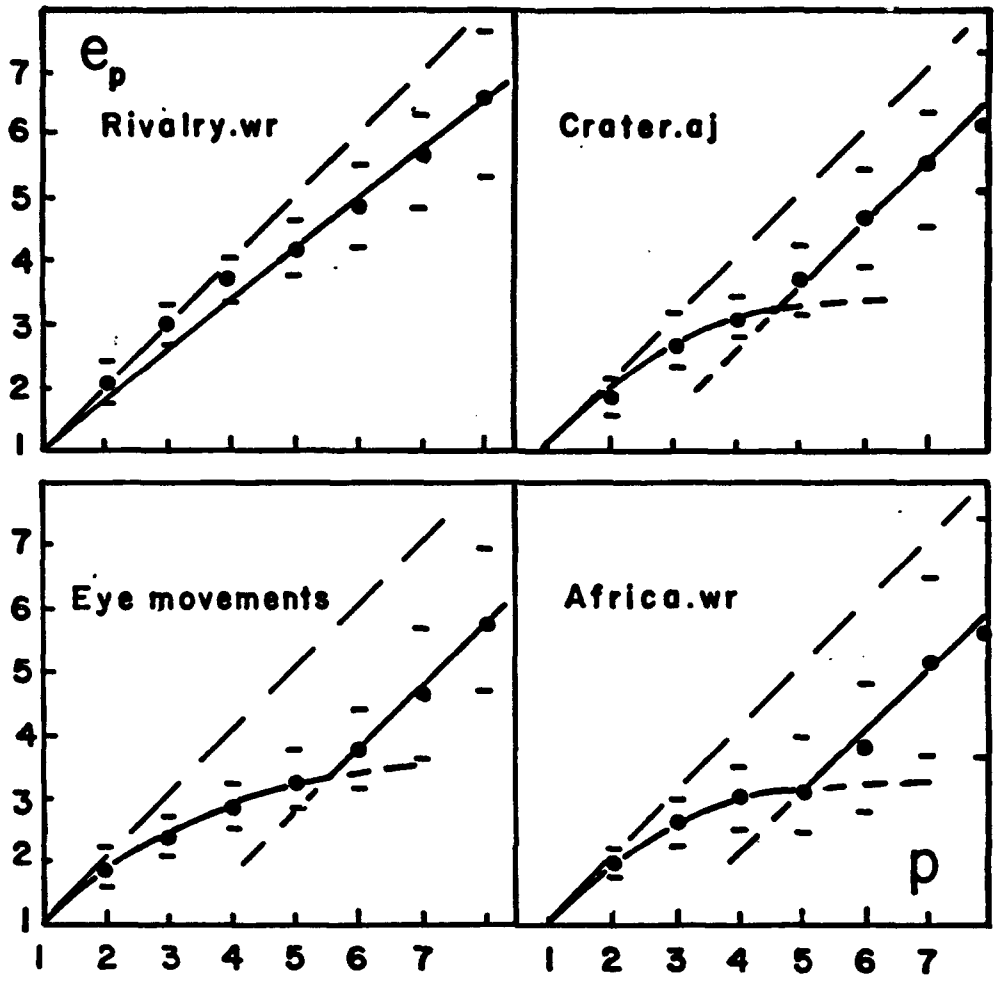



**Figure 14** Plots of $e_p$ versus the embedding dimension $p$ obtained from data like that illustrated in Figure 2. Upper left: binocular rivalry; upper right: the crater illusion of Figure 1; lower left: monkey saccadic eye movements; lower right: perceptual "eras" in "Out of Africa".

plot of $e_p$ vs $p$. This result suggests that the dominant process causing the reversals of the Necker Cube may be formally similar to the the process that leads to the competition observed during binocular rivalry.

## Eye Movements

Again using the "notch" as our indicator, evidence for chaos was also observed in the spatial pattern of 200 eye movements obtained from a monkey, who was searching a 17° × 17° field for a target (see Sommer 1993, for details). Note that in this case the analysis was not performed on the temporal intervals, but rather on the

| STIMULUS | $N$ | MEAN DURATION | DIMENSION, $p_{max}$ |
|---|---|---|---|
| Binocular Rivalry | 258 | 2.7 | – |
| Crater Illusion (aj) | 200 | 3.2 | 3.4 |
| Crater Illusion (wr) | 200 | 3.0 | 3.6 |
| Triangle and Stick | 400 | 3.2 | 3.5 |
| Duck-fish | 400 | 4.3 | 4.2 |
| Hag-girl | 200 | 2.2 | 4.8 |
| Eye Movements | 200 | 0.2 | 3.5 |
| "Out of Africa" | 320 | 25.2 | 3.3 |
| "ET" | 350 | 16.8 | 3.7 |
| "Chariots of Fire" | 350 | 19.3 | 3.5 |
| "Star Wars" | 400 | 15.2 | 3.6 |
| | | AVG. | 3.7 |

**Table 3 Number of samples $N$ and estimates of dimension underlying dynamical system, $p_{max}$. The mean durations between samples are in seconds. The eye movement analysis was performed on the spatial, not the temporal patterns of fixations; the eye movement duration reported above is the mean intersaccadic interval.**

magnitude of the saccade.[3] The location of the notch in the curve of $e_p$ vs $p$ again suggests an underlying dynamical system of dimension 3.5 which becomes masked by a stochastic process at the higher embedding dimensions. We have attempted to model these data using a combination of deterministic and stochastic processes, but to date have not succeeded. [Although the unit slope observed for $p > 5$ is suggestive of additive "random" noise, as yet we do not have a model that adequately explains all our $C(r)$ vs $(r)$ observations. The data are not typical of the expected effects of additive of noise observed in physical systems (B'en-Mizrachi et al. 1984; Theiler 1990).]

[3]The data we report were taken from the vertical component of the saccade.

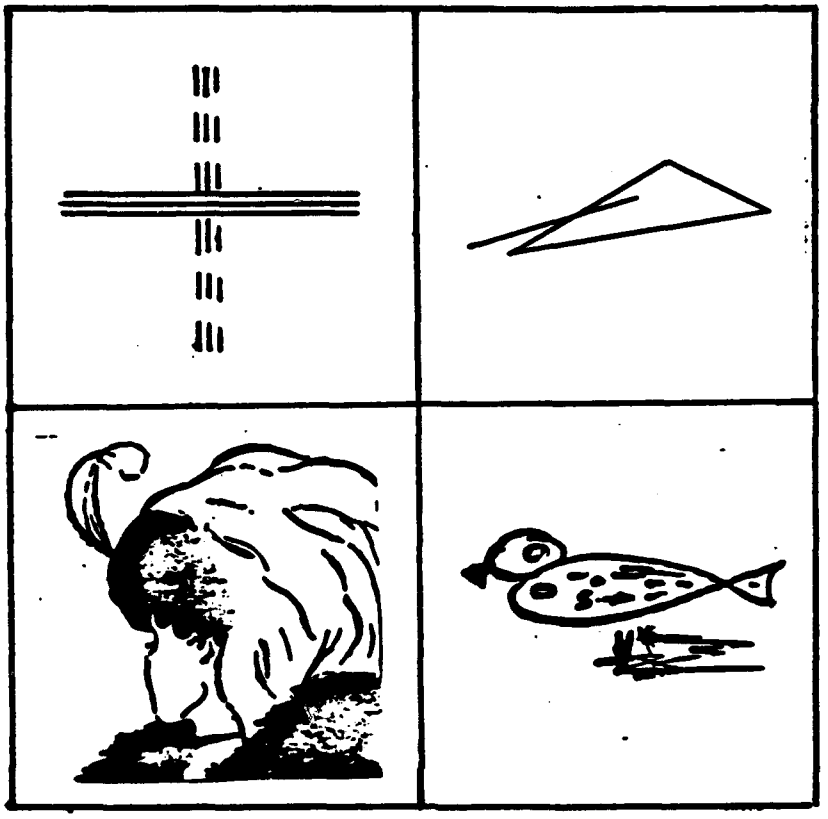


**Figure 15** Some of the stimuli used that have multiple-stable states. Upper left: a rivalrous pattern (dashed lines indicate the set presented to one eye, the solid lines indicate the set presented to the other). Upper right: triangle plus stick (stick is commonly seen as sitting on the edge of the triangle, or alternatively as angled in space above the triangle with its right end touching the interior plane); lower right: a single duck vs two fish reversal; lower left: Boring's ambiguous-mother-in-law: an old woman vs a young girl.

## Movies

The most complex perceptual activity we explored was the time sequence of segments that appear in films. Typically the camera will be directed to a person or event for a duration, then the scene flips to another person or event, as the story unfolds. Think then of the camera as the eye of the film editor, who unfolds the story, attempting to communicate the salient points. Certainly this process entails a very high level of perceptual understanding. However, the lengths of the camera segments are rather easy to measure without excessive subjective bias. For "Out of Africa", "Chariots of Fire", "Star Wars" and "ET", 200 to 400 segments were recorded. These films were chosen to represent a spectrum of fast-to-slow paced

films. The "Out of Africa" analysis is presented in the fourth panel of Figure 14. As summarized in Table 3, it can be seen that the other three films were similar, again suggesting an underlying dynamical system of dimension about 3 to 4, corrupted by stochastic noise as the embedding dimension exceeded 5. [Of interest is that "Out of Africa" also yielded an embedding dimension of less than three for the upper component of the $C(r)$ vs $(r)$ plot.][4]

## Discussion

If one views the brain as a society of competing neurons or neural processes, then their characterization as a dynamical system is not surprising. Indeed, Poston and Stewart (1978) proposed such a model for multistable percepts and such behavior has previously been reported for adaptive systems and neural activity with dimensionality estimates ranging from 3 to 7 (Altman 1991; Priesol et al. 1991; Skarda and Freeman 1987; Xu and Li 1986). It is reassuring that the observed dimensionality of their conscious counterpart – our percepts – falls in this range. What is surprising is that our estimate lies at the lower end with little variance. A dimension of less than four for any dynamical system offers a reasonable possibility for model development and study.

A second surprise is that regardless of the complexity of the perceptual task, when chaotic behavior is plausible, the inferred dimensionality remains roughly the same. It is difficult to believe that the processes underlying the reversals in the crater and the temporal sequence of segments in the movies share the same neural machinery to any great degree. Yet they both exhibit the same dimensionality! One type of process that would elicit such a result is one that is organized in a hierarchical manner having a fractal composition with roughly the same number of components collected together in each level of the hierarchy. In other words, in such a scheme, neural subsystems would compete for the attention of their "parental" systems, etc. The fractal dimension common to all tasks would then indicate that each parent has roughly the same number of "offspring". Here, we've simply adopted a proposal made elsewhere for social choice, where each state decision is not a single rational choice, but rather an aggregation of many hierarchical choice problems (D. Richards 1991). In this domain the fractal dimension also is surprisingly low.

Thirdly, with respect to the eye movement data and results, which are spatial and not temporal patterns, of interest is that again the dimensionality appears to be about 3.5. This suggests that perceptual mechanisms may be directing the spatial pattern of eye movements during the search for a target's appearance. Note that this "chaotic" search occurs even though the intersaccadic duration intervals are short, around 200 ms (Sommer 1993). The scanning strategy, then, appears controlled by a non-random, dynamical spatial representation, but not a temporal one. Again, a

[4]The other film data also suggested a lower dimensional component for values of $C(r) > .03$, but this observation remains tentative.

hierarchical search pattern, where small and large step sizes are nested as in a tree or branching structure, would be one strategy that could exhibit such deterministic chaotic behavior.

Finally, although our percepts appear to exhibit some properties consistent with deterministic chaos, this should not be construed to imply that the percepts themselves are chaotic! Indeed, the percepts can be quite rational and stable. However the state changes underlying the development of these percepts seem characteristic of a special kind of dynamical system that is well suited for searching rapidly through a large data base of variable resolution, where many states are to be evaluated quickly, probably in parallel.

## 7.0 Spinoffs

### 7.1 Texture Curvature (with Hugh Wilson)

This study examined curvature discrimination for edges created by texture contours, and includes a model incorporating end-stopped complex cells. It appears as "Curvature and separation discrimination at texture boundaries", *Jrl. Opt. Soc. Am. A*, 9: 1653-1661 (1992).

### 7.2 Shading and Stereo (Dawson & Shashua)

Pseudo stereopsis is when the binocular disparities of a surface, such as a face, are reversed but the shading is not. The impression is that the face is "normal" – the nose, for example, still points outward to the viewer.

We have manipulated noses using graphics techniques in order to push them inward, "into the head" so to speak, without altering the shading. No one is able to see these noses "shoved in". Our analysis suggests that this failure of stereopsis is simply due to the shape-from-shading solution "overriding" (in the Percepts Lattice sense) the weak stereo signal created by shaded rather than sharp contours. The effect is not special to faces, and occurs also for "playdo" shapes.

### 7.3 Configuration Stereopsis (Richards)

We are just winding up a study on 3D shape that relates to how "top-down" information about fixation distance (or shape) modulates angular disparity. Because binocular disparity appears to be computed in V2, this modulation *must* occur early in the visual pathway and hence is potentially accessible to psychophysical probing.

As the distance to an object increases, the angular disparity needed to measure the actual 3D configuration must decrease (reaching zero at the horizon). However,

if we take an object, say a cup, and evaluate its 3D shape nearby versus far away, the cup does not appear to flatten, although the disparity signal becomes much smaller as the distance increases. This suggests a rescaling of disparity with object (or fixation) distance.

We have conducted parametric studies of 3D shape from stereo over a wide range of fixation distances. The data show that indeed, the depth measure associated with a fixed angular disparity changes with fixation distance. The effect is in the direction needed to preserve the shape of 3D configurations as their distance changes, and is roughly two-thirds of what is needed for a full correction. This is evidence for neural signals being modified at or before the extraction of binocular disparity. Hence we have a preliminary "handle" on how a simple case of "bottom-up" information – namely binocular disparity – may incorporate a form of "top-down" knowledge.

### 7.4 A Neural Proposal (Ullman)

Ullman (1992) has proposed a network heirarchy scheme for how "bottom-up" information comes into register with "top-down" models. The basic process, termed "sequence-seeking", is a search for a sequence of mappings or transformations linking a source and target representation. The search is bidirectional throughout the heirarchy – "bottom-up" as well as "top-down". The novel part of the proposal is that the two searches are performed along two separate, complementary pathways, one ascending, the other descending. When a matching pattern is found, regardless of the level, then a chain of activity linking the source and target is generated, facilitating one particular path in the network. The proposal is largely consistent with what is known about cortical machinery, specifically the interplay between the various visual areas, and hence is a hypothesis about the basic scheme of information processing in the neocortex (and thalamus). Experiments related to this proposal are currently underway.

### 8.0 Publications (Supported under grant.)

Bennett, B., Hoffman, D.D. & Richards, W. (1991) Reasoning under uncertainty: Lebesgue logic and order logic. *University of California, Irvine, Math. Beh. Sci. Memo MBS 91-08.*

Feldman, J., Jepson & Richards, W. (1992) Is perception for real? *Proc. Conference on Cognition and Representation, Center for Cognitive Science Report 92-12, SUNY Buffalo, November 1992*, pp. 240-267.

Feldman, J. (1992) Constructing perceptual categories. *Proc. IEEE Computer Soc. Conf. Computer Vision & Pattern Recognition*, pp. 244-250.

Feldman, J., Epstein, D. & Richards, W. (1992) Force dynamics of tempo change in music. *Music Perception*, 10(2):185-204.

Feldman, J. (1991) Perceptual simplicity and modes of structural generation. *Proc. 13th Ann. Cog. Soc.*, pp. 299-304.

Jepson. A. & Richards, W. (1992) A lattice framework for integrating vision modules. *IEEE Trans. on Systems, Man, & Cybernetics*, 22:1087-1096.

Jepson, A. & Richards, W. (1991) What is a percept? *Occasional Paper #43, Center for Cognitive Science*, M.I.T.

Knill, D. & Richards, W. (Eds.) (1994) *Bayesian Approaches to Perception*, Cambridge University Press, in press. [Chatham Workshop, January 1993.]

Koenderink, J.J. & Richards, W. (1994) Trajectory Mapping (TM): A new scaling technique. *Proc. European Conference on Visual Perception, ECVP '93, Edinburgh, August, 1993*, in press.

Koenderink, J.J. & Richards, W. (1992) Why is snow so bright? *J. Opt. Soc. Amer. A*, 9: 645-648.

Moses, Y., Schechtman, G. & Ullman, S. (1990) Self-calibrated collinearity detector. *Biol. Cybern.*, 63: 463-475.

Richards, W., Jepson, A. & Feldman, J. (1994) Priors, preferences and categorical percepts To appear in *Bayesian Approaches to Perception*, D. Knill & W. Richards (Eds.) Cambridge University Press, in press.

Richards, W., Wilson, H.R. & Sommer, M.A. (1994) Chaos in percepts? *Biol. Cybernetics*, in press.

Richards, W. & Jepson, A. (1993) What makes a good feature? In L. Harris & M. Jenkin (Eds.) *Spatial Vision in Humans and Robots*, Cambridge University Press. Also *MIT Center for Biological Information Processing Paper No. 72* (1992) and *MIT AI Memo 1356* (1992).

Richards, W. & Jepson, A. (1993). What is a percept? *University of Toronto, Dept. Computer Science, Tech. Report RBCV-TR-93-43.*

Richards, W., Feldman, J. & Jepson, A. (1992) From features to perceptual categories.a *Proc. British Machine Vision Conference, Leeds, September 1992*, pp. 99-108.

Subirana-Vilanova, B. & Richards, W. (1991) Perceptual organization, figure-ground, attention and saliency. *MIT AI Memo 1218.*

Ullman, S. (1992) Sequence seeking and counter streams. *MIT AI Memo 1311.*

Wilson, H. & Richards, W. (1992) Curvature and separation discrimination at texture boundaries. *Jrl. Opt. Soc. Am. A*, 9:1653-1661.

Wilson, H. & Richards, W. (1989) Mechanisms of curvature discrimination (with H.R. Wilson). *J. Opt. Soc. Amer. A*, 6:106-115 (1989).